

Unidad 7

OPTIMIZANDO UNA BASE DE DATOS

Prof. Mg. Ing. Roxana Martínez
Prof. Ing. Horacio José Caballero



UAIOnline
Ultra»»



INTRODUCCIÓN A MEJORAS EN EL RENDIMIENTO DE UNA BASE DE DATOS

Unidad 7

■ OBJETIVOS

Desnormalización de una base de datos. Cómo funciona el optimizador de consultas. Optimización basada en costos. Proyección, selección y condiciones de reunión. Interpretación gráfica de los resultados del optimizador de consultas.



DESNORMALIZACIÓN

- La desnormalización es el proceso de procurar optimizar el desempeño de una base de datos por medio de agregar datos redundantes.
- A veces es necesaria porque las actuales DBMSs implementan el modelo relacional pobremente.
- Una verdadera DBMS relacional debe permitir una base de datos completamente normalizada a nivel lógico, mientras proporciona el almacenamiento físico de los datos afinado para alto rendimiento.

DESNORMALIZACIÓN

- Creando reglas en la base de datos llamadas restricciones, que especifican cómo las copias redundantes de información se deben mantener sincronizadas.
- Es el aumento en la complejidad lógica del diseño de la base de datos y la complejidad añadida de las restricciones adicionales que hacen a este acercamiento peligroso. Por otra parte, debido a los gastos indirectos de evaluación de restricciones al insertar, actualizar, o eliminar datos, una base de datos desnormalizada puede realmente ofrecer un desempeño peor que sus funcionalmente equivalentes contrapartes normalizadas. Cuando se está seleccionado o leyendo datos a menudo el desempeño será mejor.

DESNORMALIZACIÓN

Tener en cuenta que...

- Un modelo de datos desnormalizado no es lo mismo que un modelo de datos que no ha sido normalizado, y la desnormalización debe tomar lugar solamente después de que haya ocurrido un nivel satisfactorio de normalización y de que hayan sido creadas las restricciones y/o reglas requeridas para ocuparse de las anomalías inherentes en el diseño.
- Ejemplo: Que todas las relaciones estén en la tercera forma normal y cualquier relación con dependencias de unión (join) y multi-valor sean manejadas apropiadamente.

BUENAS PRÁCTICAS PARA DESNORMALIZACIÓN

- En materia de desnormalización, es aconsejable adoptar un enfoque práctico, teniendo en cuenta las limitaciones de los lenguajes como SQL y de MySql en particular, y ser prudente no desnormalizando de manera innecesaria.
- Si el rendimiento no resultara aceptable, asegurarse de comprobar si el proceso de desnormalización lo convierte en aceptable. Es aconsejable buscar alternativas, como la elección de mejor hardware para evitar la desnormalización. Resulta difícil deshacer los cambios estructurales posteriormente.

BUENAS PRÁCTICAS PARA DESNORMALIZACIÓN

- En materia de desnormalización, es aconsejable adoptar un enfoque práctico, teniendo en cuenta las limitaciones de los lenguajes como SQL y de MySql en particular, y ser prudente no desnormalizando de manera innecesaria.
- Si el rendimiento no resultara aceptable, asegurarse de comprobar si el proceso de desnormalización lo convierte en aceptable. Es aconsejable buscar alternativas, como la elección de mejor hardware para evitar la desnormalización. Resulta difícil deshacer los cambios estructurales posteriormente.

BUENAS PRÁCTICAS PARA DESNORMALIZACIÓN

- Asegurarse de que se prefiere una menor integridad de los datos a cambio de un mejor rendimiento.
- Considerar posible escenarios futuros, en los que las aplicaciones puede que planteen exigencias diferentes a los datos. El uso de la desnormalización para mejorar el rendimiento de una aplicación obliga a la estructura de datos a depender de dicha aplicación, cuando la situación ideal sería la contraria.

CUANDO APLICAR DESNORMALIZACIÓN

- Rendimiento Pobre en una consulta de varias
- tablas (con muchos datos, Ej. Millones de Registros) y realizando joins entre ellas, especialmente si es una consulta muy recurrida en nuestra aplicación y que necesita de tiempos de ejecución muy muy rápidos.

CUANDO APLICAR DESNORMALIZACIÓN

Mantenimiento de Histórico:

- Se quieren mantener en tablas de procesos los datos tal cual estaban en el momento de su creación, independientemente que en las tablas maestras hayan variado.
- Ej.: Disponemos de una tabla maestro “clientes”, que tienen un campo “domicilio” entre otros. Cada vez que un cliente nos realiza un pedido, creamos un nuevo registro en la tabla de proceso “pedidos”.

DESNORMALIZACIÓN: CONCLUSIONES

- La Desnormalización nos ayuda a mejorar el rendimiento.
- Menor número de tablas.
- Redundancia de Datos.
- Mejora Velocidad de las consultas.
- Menor Integridad de datos.



CONCEPTO - ASPECTOS DE LA OPTIMIZACIÓN DE LAS CONSULTAS

El **optimizador** de consultas es el componente del sistema de gestión de base de datos que intenta determinar la forma más eficiente de ejecutar una consulta SQL, es decir, cual es, de los posibles planes de ejecución para una consulta dada, el más eficiente.

CONCEPTO - ASPECTOS DE LA OPTIMIZACIÓN DE LAS CONSULTAS

- Conversión a ecuaciones equivalentes del Álgebra Relacional y selección de la más eficiente. (Procedimiento de fuerza bruta)
- Estrategias para el procesamiento de la consulta basadas en:
 - Heurística
 - Estadísticas
 - Índices
 - Vistas materializadas

(Todos estos son procedimientos heurísticos o inteligentes para disminuir el costo de cálculo de planes alternativos)

EJEMPLO – CONSULTA BÁSICA

- La expresión para un banco en SQL:

“Encontrar todos los clientes que tengan cuenta en cualquier sucursal de Capital Federal”

- La expresión en álgebra relacional sería:

$\Pi \text{ NombreCliente } (\sigma \text{ CiudadSucursal} = \text{"Capital Federal"} (\text{Sucursal} \bowtie (\text{Cuenta} \bowtie \text{Cliente})))$

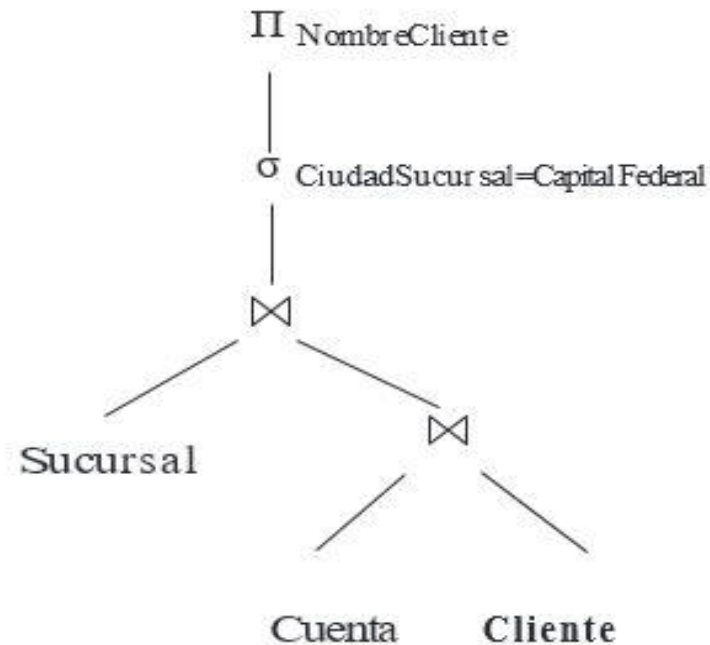
- Una expresión equivalente mucho más eficiente sería:

$\Pi \text{ NombreCliente } ((\sigma \text{ CiudadSucursal} = \text{"Capital Federal"} (\text{Sucursal})) \bowtie (\text{Cuenta} \bowtie \text{Cliente}))$

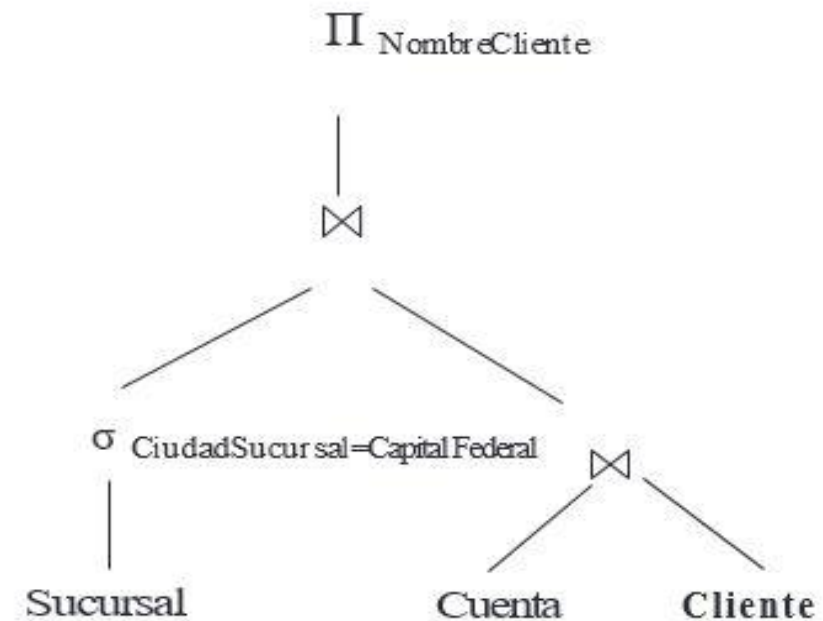
|

EJEMPLO – CONSULTA BÁSICA

- Gráficamente en árbol:



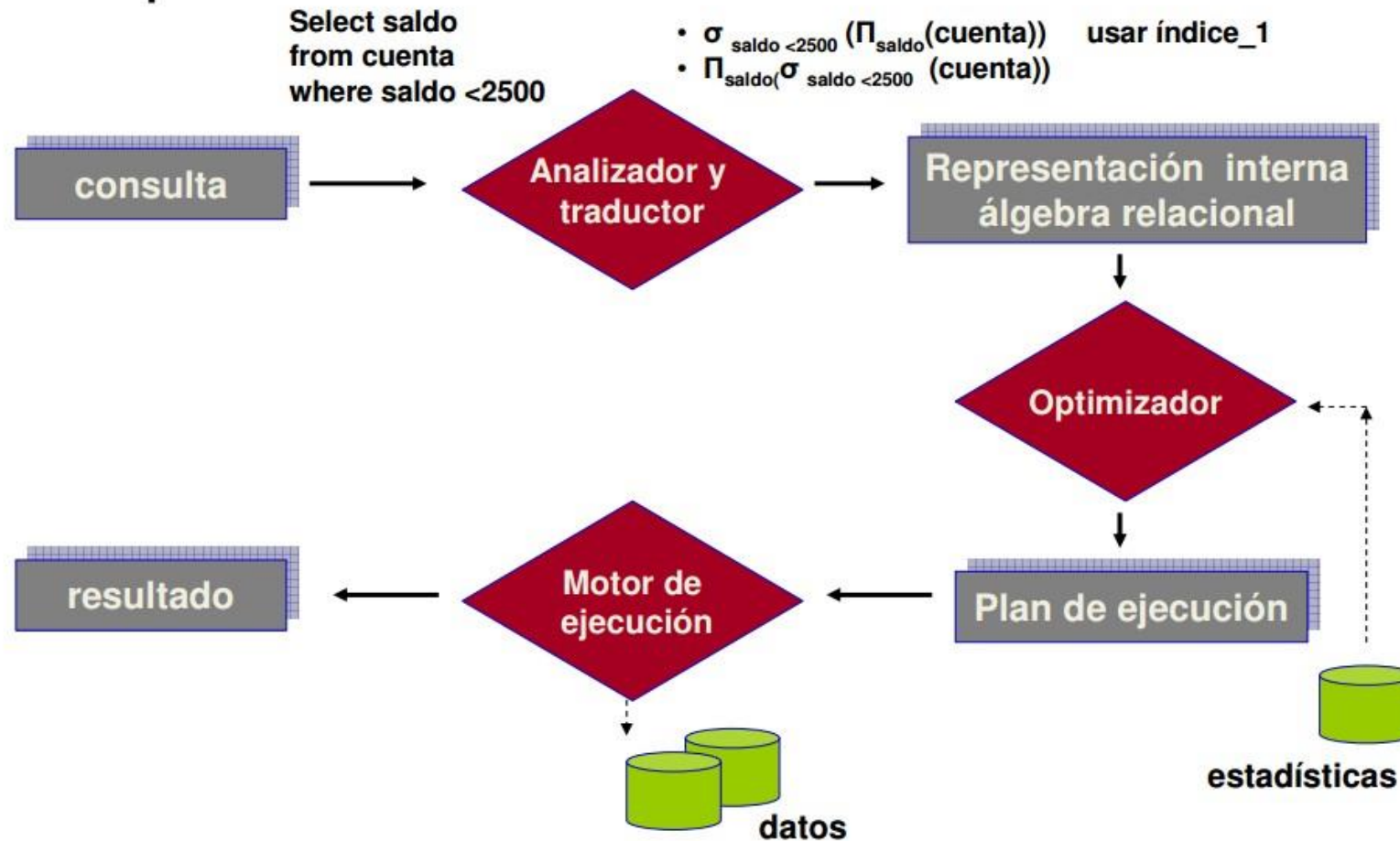
(a) Árbol inicial de la expresión



(b) Árbol transformado de la expresión

Expresiones equivalentes.

CONCEPTO – PLAN DE EJECUCIÓN - ETAPAS



CONCEPTO – PLAN DE EJECUCIÓN - ETAPAS

- La generación del plan de ejecución sigue las siguientes etapas:
 - La generación de expresiones que sean lógicamente equivalentes a la expresión dada.
 - La estimación del costo de cada expresión.
 - El registro de las expresiones alternativas resultantes y sus combinaciones para generar planes distintos de ejecución.
 - La estimación del costo de las combinaciones propuestas.
 - La elección del mejor plan de ejecución.

HEURÍSTICAS

■ Heurísticas de Optimización

Un inconveniente de la optimización basada en el coste es el coste de la propia optimización. Por ello, muchos sistemas usan la heurística para reducir el coste de la optimización.

- Realizar las operaciones de selección tan pronto como sea posible.
- Realizar las proyecciones tan pronto como sea posible.
- Realizar los productos cartesianos tan tarde como sea posible.

ESTADÍSTICAS

- Para que puedan basar su elección de estrategia en información de confianza, los sistemas de bases de datos almacenan estadísticas para cada relación y de las ejecuciones de las consultas. Entre estas estadísticas están
- El número de tuplas de la relación r .
- El tamaño del registro (tupla) de la relación r en bytes.
- El número de valores diferentes que aparecen en la relación r para un atributo determinado.
- Los tiempos de ejecución de las consultas en distintos horarios.

ÍNDICES

- Con índices sobre campos que son condiciones de selección, ordenamiento o agrupamiento se pueden acelerar las consultas.
- Por lo tanto se analiza estadísticamente a esos campos para determinar si se necesita indexarlos y qué tipo de índice aplicar.

VISTAS MATERIALIZADAS

- Las vistas materializadas pueden usarse para acelerar el procesamiento de las consultas.
- La conservación incremental de las vistas es necesaria para actualizar de forma eficiente las vistas materializadas cuando se modifican las relaciones subyacentes.
- El diferencial de cada operación puede calcularse mediante expresiones algebraicas que impliquen a los diferenciales de las entradas de la operación.
- Entre otros aspectos relacionados con las vistas materializadas están el modo de optimizar las consultas haciendo uso de las vistas materializadas disponibles y el modo de seleccionar las vistas que hay que materializar..



¿Consultas?



UAI

**Universidad Abierta
Interamericana**